

Spatial variation and prediction of forest biomass in a heterogeneous landscape

S. Lamsal • D. M. Rizzo • R. K. Meentemeyer

Received: 2011-02-18; Accepted: 2011-08-12

© Northeast Forestry University and Springer-Verlag Berlin Heidelberg 2012

Abstract: Large areas assessments of forest biomass distribution are a challenge in heterogeneous landscapes, where variations in tree growth and species composition occur over short distances. In this study, we use statistical and geospatial modeling on densely sampled forest biomass data to analyze the relative importance of ecological and physiographic variables as determinants of spatial variation of forest biomass in the environmentally heterogeneous region of the Big Sur, California. We estimated biomass in 280 forest plots (one plot per 2.85 km²) and measured an array of ecological (vegetation community type, distance to edge, amount of surrounding non-forest vegetation, soil properties, fire history) and physiographic drivers (elevation, potential soil moisture and solar radiation, proximity to the coast) of tree growth at each plot location. Our geostatistical analyses revealed that biomass distribution is spatially structured and autocorrelated up to 3.1 km. Regression tree (RT) models showed that both physiographic and ecological factors influenced biomass distribution. Across randomly selected sample densities (sample size 112 to 280), ecological effects of vegetation community type and distance to forest edge, and physiographic effects of elevation, potential soil moisture and solar radiation were the most consistent predictors of biomass. Topographic moisture index and potential solar radiation had a

positive effect on biomass, indicating the importance of topographically-mediated energy and moisture on plant growth and biomass accumulation. RT model explained 35% of the variation in biomass and spatially autocorrelated variation were retained in regression residuals. Regression kriging model, developed from RT combined with kriging of regression residuals, was used to map biomass across the Big Sur. This study demonstrates how statistical and geospatial modeling can be used to discriminate the relative importance of physiographic and ecologic effects on forest biomass and develop spatial models to predict and map biomass distribution across a heterogeneous landscape.

Keywords: forest biomass; landscape heterogeneity; spatial variation; semivariogram; regression tree; regression kriging; Big Sur California

Introduction

Forest biomass holds more carbon than any other terrestrial biome on Earth (Houghton 1999). Scientists have recently recognized that the spatial distribution of forest biomass is critical to measuring trends in forest carbon stocks through time (Hu and Wang 2008), predicting the global carbon cycle (Houghton 2005), and mapping the risk of wildfire (Mickler et al. 2002; He et al. 2004). Although our knowledge of how deforestation patterns influence large-area estimates of carbon storage has increased, especially in the tropics (e.g. Defries 2002), assessment of forest biomass distribution in spatially heterogeneous landscapes continues to present challenges especially in regions with insufficient density of forest inventory plots. Across highly heterogeneous landscapes, spatial variation in forest biomass may be especially complex where variations in plant growth and species composition occur over short distances due to sharp interactions between physiographic (e.g., elevation and solar radiation), ecological (e.g., forest edge effects and fire history), and human-mediated (e.g., past logging and other land-use changes) factors (Anderson et al 2009; Clark and Clark 2000; Poulos 2009). As environments become increasingly complex, approaches that examine landscape heterogeneity and its influence on forest biomass distribution are greatly needed to accurately assess trends in the global carbon balance (Houghton 2005).

Foundation project: This work was financially supported by the National Science Foundation (EF-0622770 and EF-0622677), the USDA Forest Service – Pacific Southwest Research Station, and the Gordon & Betty Moore Foundation.

The online version is available at <http://www.springerlink.com>

S. Lamsal (✉) • R. K. Meentemeyer
Department of Geography and Earth Sciences, University of North Carolina, Charlotte, NC, U. S. A.
Phone (+1)704-687-5922; Fax (+1)-704-687-5966 ;
Email sanjaylamsal@gmail.com (S. Lamsal); rkmeente@uncc.edu (R. K. Meentemeyer)

D. M. Rizzo
Department of Plant Pathology, University of California, Davis, CA,
U. S. A. Email: dmrizzo@ucdavis.edu

Responsible editor: Chai Ruihai

Regional scale assessments of forest biomass distribution have mostly relied on geographically extensive forest inventory data. In the United States, for example, the Forest Inventory and Analysis program (FIA; Smith 2002) provides the most practical data for estimating forest biomass at regional to national scales (Brown et al. 1999; Blackard et al. 2008; Hu and Wang 2008). However, with an average density of one plot per ~2400 ha, FIA data may be too sparsely distributed for many landscape-scale studies of biomass variation. In a recent study, Freeman and Moisen (2007) observed highly erratic patterns of spatial variation in biomass distribution derived from FIA data for regions in the western North America. To counter such problem, models of biomass distribution often aggregate biomass estimates to coarser regions, such as political boundaries (e.g., Meng et al. 2007). Alternatively, biomass distribution models can be based on correlations with key environmental variables that may be applied in a geographic information system (GIS) to map spatially-explicit patterns of biomass (Bacinni et al. 2004; Blackard et al. 2008). As spatial data layers of environmental and ecological variables are becoming increasingly available at finer spatial resolution and across larger spatial extents, correlation-based models offer prospects for developing finer resolution estimates and maps of biomass at regional to national extents.

The processes that underly plant growth and biomass accumulation in heterogeneous landscapes are influenced by a number of ecological and physiographic factors that collectively influence the flow of energy, moisture, and nutrients available to vegetation communities (Chen et al. 1999; Lovett et al. 2005; Saatchi et al. 2009). For example, physiographic influences of topography on variations in climate, solar radiation, and soil moisture have been increasingly incorporated into models of regional plant biomass distribution (e.g. Bacinni et al. 2004). However, ecological drivers of biomass that vary at finer spatial scales have received less attention in models of biomass distribution (Vanwalleghe and Meentemeyer 2009). Ecosystem disturbances caused by wildfire, for instance, can cause abrupt transitions in landscape patterns of vegetation, which may in turn mediate biomass regrowth and accumulation (Ohmann et al. 2007). Overall, ecological and physiographic drivers may have complex interactive effects that should be evaluated to assess biomass spatial variation and develop biomass prediction models.

In this study, we use statistical and geospatial modeling to examine spatial variation in above-ground forest biomass distribution in an environmentally heterogeneous landscape. Using the Big Sur ecoregion (~80,000 ha) in coastal California as a case study of a spatially heterogeneous region, we examine the hypothesis that biomass distribution is spatially autocorrelated and their spatial patterns are related to landscape-scale variations in ecological and physiographic factors.

We estimate forest biomass in 280 field plots from redwood-tanoak and mixed evergreen forest that span across a range of environmental conditions. The density of these data provides a unique assessment of biomass spatial variation which is not feasible with sparser regional scale data. We assess spatial dependence and local uncertainty in biomass distribution using semivariogram analyses, and develop regression models to pre-

dict the effects of ecological and physiographic factors on biomass distribution. We developed geospatial models of biomass distribution based on biomass spatial autocorrelation and its relationships with ecological and physiographic landscape variables. Finally, using random sample selection, we assess the role that the spatial distribution and density of field sampling plays on the performance of biomass models in heterogeneous landscapes.

Materials and methods

Study system

The Big Sur ecoregion (79356 ha) extends along the western slope of the Santa Lucia Range, from Point Lobos south to Salmon Creek, CA (Fig 1). The region's adjacency to pacific ocean provides a mediterranean-type climatic condition with warm dry summers and mild wet winters. The topography is highly dissected by steep slopes and drainages with elevations ranging from sea level to 1571 m. This environmentally complex region supports a diversity of plant communities (Henson and Usner 1996). Upper elevation slopes and rocky ridges tend to support mixed coniferous forests composed of ponderosa pine (*Pinus ponderosa*), sugar pine (*P. lambertiana*), jeffrey pine (*P. jeffreyi*), coulter pine (*P. coulteri*) and Santa Lucia fir (*Abies bracteata*). Drier south-facing slopes and ridges at mid elevations are often dominated by chaparral shrubland and annual grasslands. Mixed evergreen forests, consisting of coast live oak (*Quercus agrifolia*), Shreve's oak (*Q. parvula* var. *shrevei*), bay laurel (*Umbellularia californica*), and madrone (*Arbutus menziesii*), typically occur on moister slopes which transition to riparian corridors of redwood (*Sequoia sempervirens*) - tanoak (*Lithocarpus densiflorus*) forest at lower elevations. Low elevation south- and west-facing slopes support drought deciduous coastal sage scrub vegetation (Borchert et al. 2004). Spatial structure of the vegetation communities in the Big Sur is thought to be largely determined by disturbance and physical environment (Davis et al. 2010). Fire is the major ecosystem disturbance in the Big Sur and its interactive effects with physical environment has influenced the direction and rate of changes in vegetation communities (Callaway and Davis 1993). Further, the rugged topography of the region mediates drought severity which influences the patterns of seedling recruitment type, and species that depend on fires for recruitment are favored by increasing drought severity (Meentemeyer and Moody 2002). Proximity to the Pacific Ocean modifies temperature year round providing a mild climate with minimal variations. Based on Parameter-elevation Regressions on Independent Slopes Model (PRISM, Daly et al. 2001), the region has a long term average annual precipitation of 40.5–145.0 cm, minimum temperature of 5.5–9.6 °C and maximum temperature of 16.9–21.3 °C.

Field plot data

Over the summers of 2006 and 2007, we established 280 randomly distributed field plots (500 m²) in mixed evergreen (n =

162) and redwood-tanoak ($n = 118$) forest communities (Fig. 1). We recorded the plot center using GPS (global positioning system) receivers with a horizontal accuracy within 1 m using differential correction (Trimble Navigation Limited, Sunnyvale, CA) and mapped the location of every live and dead stem ≥ 1 cm DBH (diameter at breast height, 1.3 m) relative to the plot center. For each tree, we identified the species and measured its DBH. We estimated the above-ground biomass of individual live trees greater than 2.5 cm DBH using allometric equations summarized by Jenkins et al (2003). The DBH measurements were taken from 8812 trees and DBH distribution had a mean of 18 cm, minimum 2.5 cm, maximum 243.7 cm, and standard deviation 24.4 cm. The estimated biomass of all trees within a plot were summed to represent plot biomass (expressed in Mg/500 m²). The distance between a plot and its closest neighbor ranged from 164 m to 2990 m with 650 m as the average distance between a plot and its closest neighbor.

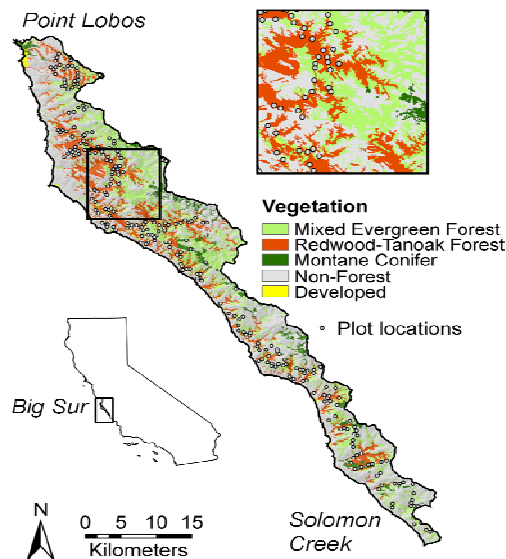


Fig 1. Study system showing vegetation types across the Big Sur ecoregion and plot locations. The inset shows a detail about spatial heterogeneity in vegetation communities and spatial sampling design.

Ecological and physiographic variables

Within a GIS, we assembled raster maps of ecological and physiographic variables (Table 1) that we hypothesized to influence plant growth and biomass accumulation in the study region. The spatial distribution of mixed evergreen, redwood-tanoak forest, montane conifer and other forested vegetation types was mapped by Meentemeyer et al. (2008) using a combination of high resolution imagery and field data (Fig 1). We quantified the heterogeneity in vegetation types surrounding each plot as the proportion of forest and non-forest vegetation type at varying radii (50, 100, and 200 m) and measured the distance of each plot to the nearest forest edge. To assess landscape scale variation in soil properties, we used maps of soil orders and sub-orders from the Soil Survey Geographic (SSURGO) database (<http://soildatamart.nrcs.usda.gov> accessed March 2009). We identified the number of fire events since 1900, 1950 and 1980 using intraagency fire history data from co-operative efforts of the California Department of Forestry and Fire Protection, United States Department of Agriculture Forest Service Region 5, Bureau of Land Management, and National Park Service, and distributed by the United States Forest Service (<http://www.fs.fed.us/r5/rs/clearinghouse/>).

To assess physiographic variations, we mapped elevation, slope, topographic moisture index (TMI), and potential solar radiation (PSR) for the months of March, June and December using a 30 m digital elevation model (DEM) (Table 1). TMI characterizes topographic redistribution of soil moisture and is computed as natural log of the ratio between upslope drainage area and local slope gradient (Moore et al. 1991). PSR characterizes topographic variations in potential incoming solar radiation using cosine of illumination on slope equation (Dubayah 1994). Climatic variation was mapped using 30-year average annual precipitation, and minimum, maximum, and mean temperature. We also mapped latitude and proximity to the Pacific Ocean as indirect gradients of geographical and physiographical effects on climate. Values for each of the nine variables were identified at all 280 plot locations.

Table 1. Ecological and physiographic variables assembled for the Big Sur region

Variables	Data source
Vegetation community type – Mixed Evergreen (ME) and Redwood-tanoak (RWTO)	Vegetation map by Meentemeyer et al (2008)
Distance to forest (Dist_For) and non-forest (Dist_NonFor) in meter	Vegetation map by Meentemeyer et al (2008)
Soil order (Ord) and sub-order (Sub_Ord)	Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture
Number of fire outbreaks (N)	California Department of Forestry and Fire protection, United States Forest Service, Bureau of Land Management, and National Park Service.
Elevation in meter (E), slope in degree (S), topographic moisture index (TMI), and potential solar radiation index (PSR) for the month of March (PSR_Mar), June (PSR_June) and December (PSR_Dec)	United States Geological Survey, National Elevation Dataset (Gesch, D.B., 2007).
Temperature in °C (T) and precipitation in meter (P)	PRISM (Daly et al. 2001)
Distance to coast in meter (Dist_Coast)	County boundary map, California Department of Conservation
Latitude in meter (X)	Study area boundary

Spatial autocorrelation and local uncertainty

We assessed spatial dependence in biomass distribution between field plots using semivariance analyses (Webster and Oliver 2007). Our plot data were normalized for semivariance analyses using natural log transformation. Semivariance is a measure of dissimilarity of observations across space and was computed from our sample data as:

$$\hat{\gamma}(h) = \frac{1}{2n} \sum_{i=1}^n \{z(x_i) - z(x_i + h)\}^2 \quad (1)$$

where, n is the number of pairs of sample observations separated by a distance h , and $z(x_i)$, $z(x_i + h)$ are sample measurements separated by a distance h . The plot of $\hat{\gamma}(h)$ against h is known as a semivariogram. We used the semivariogram to summarize spatial variation as a function of lag separation using three parameters – nugget, sill and range. The relative nugget effect, computed as the ratio of nugget to sill, was used as an index of spatially correlated variation.

We modeled the local uncertainty in biomass distribution using a conditional cumulative distribution function (ccdf), which summarizes probability that the value of a variable is below a certain value, conditioned to the observed values of the variable. This non-parametric approach to assessing local uncertainty of a spatially distributed variable transforms the observed values into a vector of K indicators consisting of 0 and 1 using the transformation function $I(x; z_c)$ (Goovaerts 1997).

$$I(x; z_c) = \begin{cases} 1, & z(x) \leq z_c \\ 0 & \text{otherwise} \end{cases} \quad c = 1, \dots, K \quad | (n) \quad (2)$$

where, z_c is a threshold value, K is a user-specified number of cutoff thresholds, and $| (n)$ indicates conditioning to local information, i.e., data in neighboring locations. We quantified spatial structure of local uncertainty using semivariograms of indicator transforms, often called indicator semivariograms. In order to model local uncertainty in biomass distribution, we transformed biomass data into 1 and 0 based on seven threshold values from its histogram (second to eighth deciles, which are 8.59, 10.03, 13.19, 15.46, 19.53, 22.93, and 29.18) and computed the indicator semivariogram. We excluded the first and ninth deciles because indicator semivariograms of extreme thresholds are not well defined.

Biomass prediction models

Spatial autocorrelation in biomass distribution and its relationship with ecological and physiographic variables allowed us to develop two types of biomass prediction models: (1) geostatistical prediction model that rely on spatial autocorrelation, popu-

larly known as kriging (Webster and Oliver 2007), and (2) environmental correlation model (McKenzie and Ryan 1999) based on the predictive relationship of biomass with ecological and physiographic variables. To take advantage of both prediction methods, we developed a hybrid prediction model, popularly known as Regression Kriging (RK), which combines deterministic spatial variation described by a regression model with geostatistical prediction of regression residuals using kriging (Hengl et al. 2004; Vanwalleghe and Meentemeyer 2009). RK assumes that prediction of a variable $\hat{Z}(x_i)$ at an unvisited location (x_i) is an additive function of variables describing spatial and environmental variation:

$$\hat{Z}(x_i) = m(x_i) + \mathcal{E}(x_i) + \mathcal{E}' \quad (3)$$

where, $m(x_i)$ is a structural component that can have a constant mean or exhibit a trend, $\mathcal{E}(x_i)$ is the random but spatially autocorrelated variation, and \mathcal{E}' is the spatially uncorrelated residual error term.

To model deterministic spatial variation, we explored predictive association of biomass with ecological and physiographic variables using regression tree (RT) analyses. RT is a non-parametric tree-based regression approach that follows binary recursive partitioning whereby tree models are constructed by repeated splitting of the set of observations (parents nodes) into two descendent subsets (child nodes) such that data within child nodes are relatively homogeneous (Breiman et al. 1984). It is ideally suited for the analysis of complex ecological data which requires flexible and robust analytical methods to deal with nonlinear relationships and higher order interactions (DeAth and Fabricius 2000). We built several tree models of biomass variation using 10-fold cross validation (Breiman et al 1998) with $v = 10$ in our case, and the model that explained the highest amount of variation during validation (cross validation R^2) was selected as the optimal prediction model. In v -fold cross validation, the data are divided into v subsets of approximately equal size, the model is trained v times leaving out one of the subsets from training and using this omitted set for model validation. To model autocorrelated spatial variation, we assessed spatial autocorrelation in RT prediction residuals using semivariogram, and mapped regression residuals using ordinary kriging (OK). RT and OK differ in the sense that RT is an aspatial approach based on the predictive associations of biomass with environmental covariates, while OK is a univariate spatial approach that relies on the model of spatial autocorrelation structure and observed data to derive spatial interpolations. Finally, RT model was combined with OK of regression residuals to develop RK model.

To assess model prediction quality, we randomly split the data into model (75% data, $N = 210$) and validation set (25%, $N = 70$); prediction models were developed using model set and their predictions were compared at validation locations. We assessed overall model fit using coefficient of determination (R^2) and model prediction quality using the Root Mean Squared Error

(RMSE). The R^2 indicates an overall agreement between predicted and true values, while RMSE indicates overall prediction quality by comparing predicted and true values.

Effect of sample size

Accurate predictions of spatially autocorrelated variables using RK depends on the performance of environmental correlation models, which largely depends on sample size. To assess the effects of sample size and selected samples on environmental correlation models, we developed RT models using varying sample sizes developed from two sampling schemes. First, we developed models from randomly selected 80%, 70%, 60%, 50%, 40%, and 30% of the data, and compared model fit using R^2 . Second, we randomly selected and set aside 20% of the data as independent validation data (N: 56) and developed models from the remaining 80%, 70%, 60%, 50%, 40%, and 30% data. This resulted in 6 model datasets consisting of 224, 196, 168, 140, 112 and 84 samples. Both sampling schemes were implemented within GIS, and regression analyses were done within a statistical package. For each sample size, the regression modeling was done 10 times by random selection of modeling data resulting in a total of 60 models. The difference in variance explained by models of different sample size was used to assess the effect of sample size on RT model, and the difference in variance explained by 10 models developed from a given sample size was used to assess the effect of samples on RT model. We also assessed spatial autocorrelation structure in biomass distribution using reduced sample sizes.

We used ArcGIS 9.2 (Environmental Systems Research Institute, Redlands, CA) for GIS analyses, spatial interpolation and map visualization, Hawth's Analyses Tool (Beyer, H. L. 2004) for random sampling to assess the effects of sample size, and JPM (SAS, Cary, NC) for regression analyses.

Results

Biomass autocorrelation and uncertainty

Biomass distribution in the Big Sur ranged from 0.8 to 139.6 Mg/500 m² with an average density of 22.6 Mg/500 m². The distribution was wide (coefficient of variation 80%) and positively skewed with density over 50 Mg/500 m² in 22 plots and over 100 Mg/500 m² in only one plot. Plots with large biomass density resulted from a few large diameter trees rather than many smaller trees within a plot. Biomass differed significantly between mixed evergreen and redwood-tanoak forests ($F = 68.8$, $p < 0.00$) with an average biomass density of 15.5 for mixed evergreen and 31.7 for redwood-tanoak forests.

Spatial autocorrelation in biomass distribution was modeled with a spherical semivariogram (Fig 2) fitted using the parameters: nugget 0.33, sill 0.59 and range 3 185 m. The range (3,185 m) suggests that biomass was spatially autocorrelated at sub-regional scale, and plots with separation distances greater than 3.1 km may be considered to be spatially independent. Indicator

semivariograms were modeled using a sum of nugget effect and spherical model with semivariograms parameters summarized in Table 2. The structure of indicator semivariograms varied along the transition from low to high thresholds and the nugget effect increased from 4% in semivariogram for the second decile to 49% for the eighth decile. The autocorrelation range increased with higher threshold values and ranged from 818 to 3 650 m, which suggests that lower biomass areas were locally constrained while higher biomass areas were spatially autocorrelated across larger distances.

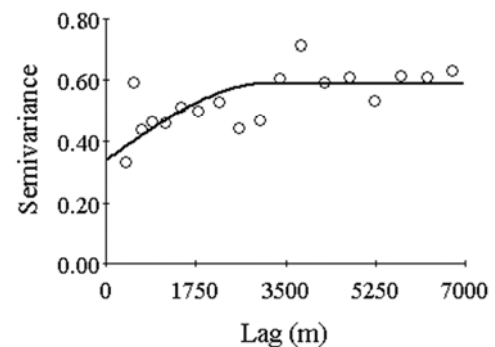


Fig 2. Semivariogram of biomass density

Table 2. Parameter used to model indicator semivariograms of biomass density

Threshold deciles	Nugget	Sill	Range	Nugget effect (%)*
Second	0.007	0.169	818	4
Third	0.012	0.208	1058	6
Fourth	0.024	0.234	1545	10
Fifth	0.052	0.253	1600	20
Sixth	0.090	0.243	2245	37
Seventh	0.089	0.205	2850	43
Eighth	0.077	0.158	3650	49

*computed as the ratio of nugget to sill variance and expressed as %

Biomass prediction models

The optimal RT model for biomass prediction had eight prediction nodes, which were developed using six predictor variables (vegetation, TMI, northing, PSR_Dec, elevation, and distance to non-forest) (Fig 3). The model explained 35% of the variation in biomass with a cross validation RMSE of 14.8. Vegetation type was identified as the first splitting variable, with higher biomass in redwood-tanoak forests, and the effects of other predictor variables were nested within the effects of vegetation type on biomass. Geographic differences, potential solar radiation and elevation were predictors of biomass variation in mixed-evergreen forests, while TMI, distance to non-forest and elevation were predictors in redwood-tanoak forest. For both vegetation types, lower biomass were found in higher elevation plots and vice-versa. As redwood-tanoak forests are found in lower

elevation areas, a smaller threshold was selected for predictive splitting in redwood-tanoak plots (elevation = 381 m) compared to mixed-evergreen plots (elevation = 767 m). Residuals from RT predictions were spatially autocorrelated upto 2,330 m, and RK model implemented within GIS produced spatially varying

patterns of biomass distribution across the Big Sur (Fig 4). When model prediction quality were assessed at independent validation sites (N: 70), RT model explained 26% of biomass variation with prediction RMSE of 15.8, and RK model explained 41% of biomass variation with prediction RMSE of 13.2.

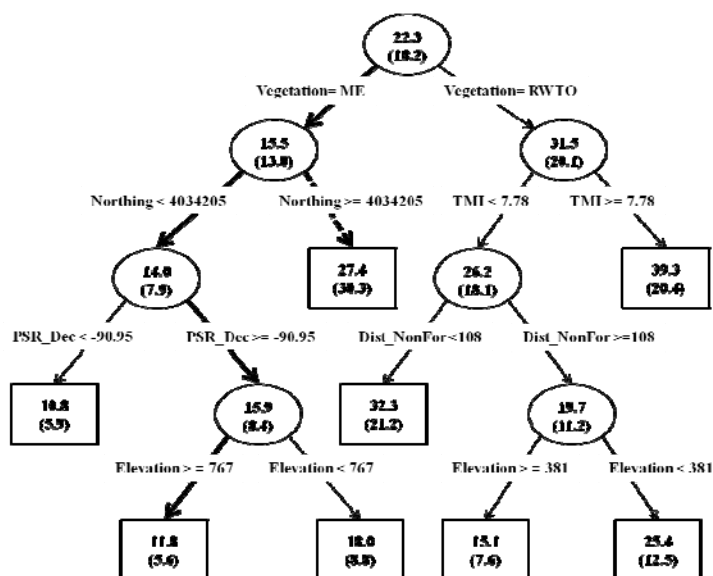


Fig 3. Predictive structure of Regression Tree model for biomass prediction based on ecological and physiographic variables. Ovals and squares represent non-terminal and terminal nodes, respectively.

The values inside the nodes are the predicted values with prediction standard deviation within the bracket. Prediction criteria are presented between the nodes. The dark solid arrows indicate the combination of environmental conditions associated with best predictions, while dashed arrow indicates the combination of environmental conditions associated with poorest prediction. Abbreviations for predictor variables are – MO: mixed oak, RWTO: redwood-tanoak, PSR_Dec: Potential solar radiation for the month of December, Dist_NonFor: distance to the non forest, TMI: topographic moisture index.

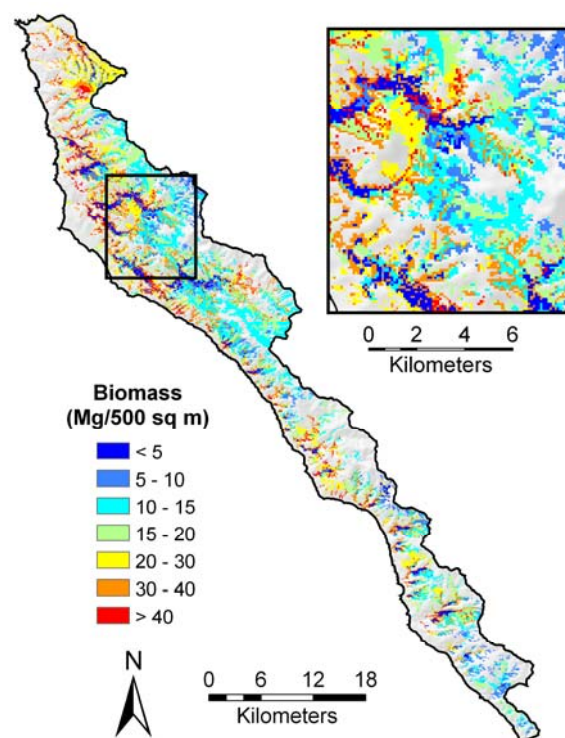


Fig 4. Biomass map derived using Regression Kriging. The inset shows detailed pattern of biomass spatial variation.

Effects of sample size

RT models developed with sample sizes ranging from 112 to 280 explained 32%–45% of the variance in biomass and the reduction in sample size did not have a significant effect on model fit (Table 3). Across all sample sizes, vegetation type and TMI were consistently identified as predictors of biomass. PSR was also selected across sample sizes except the scenario that used 112 observations. As expected, RT models based on smaller sample sizes had fewer prediction nodes, which varied from 8 nodes with 280 samples to 4 nodes with 112 samples. For a given sample size (with ten models) and across various sample sizes of 224, 196, 168, 140, 112 and 84, RT model fit did not differ significantly (Fig 5). However, when the models were tested with an independent dataset, prediction quality appeared to be poorer than as expected with the model set. When tested for spatial autocorrelation, sample sizes of 112 and lower did not exhibit spatial autocorrelation structure.

Table 3. Predictor variables, number of prediction nodes, and variance explained by regression tree prediction model developed with varying sample sizes

Sample size	Predictor variables	N nodes	R ²
280	Vegetation, TMI, PSR_Dec, Elevation, Northing*, Dist_NonFor	8	0.33
252	Vegetation, TMI, PSR_Dec, Elevation, Northing	7	0.32
224	Vegetation, TMI, PSR_Dec, Elevation, Northing, Dist_NonFor	7	0.31
196	Vegetation, TMI, PSR_Dec, Distance to non forest	4	0.45
168	Vegetation, TMI, PSR_Dec	5	0.32
140	Vegetation, TMI, PSR_Dec	4	0.35
112	Vegetation, TMI	4	0.32

Abbreviations - TMI: topographic moisture index, Dist_NonFor: distance to non-forest; PSR_Dec: potential solar radiation for the month of December

* Northing represents latitude in meter

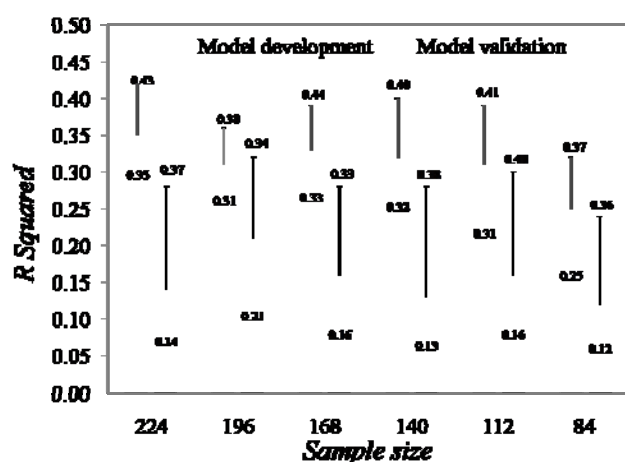


Fig 5. Effects of sample size and samples on regression tree prediction performance during model development (model fit) and validation (model validation). Size of the box indicate average R^2 which are also plotted within the box, the error bars indicate standard deviation of R^2 , and the maximum R^2 are plotted above the bars

Discussion

We assessed spatial variation of above-ground biomass distribution in mixed-evergreen and redwood-tanoak forests that extend across heterogeneous landscapes of the Big Sur ecoregion, and developed models to predict and map biomass at unsampled locations. Semivariograms revealed that biomass distribution was spatially structured and autocorrelated up to 3,185 m. Semivariance in the first lag, which was estimated from 25 sample pairs with an average spacing of 375 m, accounted for over half of the sill variance. Therefore, substantial portion of biomass spatial variation in the Big Sur are localized, which may be attributed to the effects of ecological and physiographic gradients (Blackard et al. 2008) that constitute the landscape heterogeneity of the Big Sur. This corroborates with recent study (Freeman and Moisen 2007) which observed weak spatial autocorrelation in biomass distribution for some regions in western US; such weak autocorrelation is very likely related to the scale of analyses whereby biomass distribution may be autocorrelated within spatial scales of few kilometers (as in Big Sur) that are shorter than the lag spacing (5–100 km) that was used to model biomass semivariogram. The range of indicator semivariograms increased with higher threshold values (Table 2), which suggests that spatial clusters of lower biomass areas were locally constrained and those of higher biomass areas extended across larger distances. This confers to spatial patterns of vegetation distribution that result from smaller trees being clumped within finer spatial scales, while larger trees are spatially distributed in a fairly regular pattern across larger distances so as to achieve dominance by avoiding direct competition for nutrients and sunlight from other trees of similar age (Gilbert and Lowell 1997).

Biomass spatial variation across heterogeneous landscapes of the Big Sur can be explained based on ecological and physi-

ographic variables identified as predictor variables in the RT models. The RT methodology uses a series of binary splits to develop prediction models that are suited for a subset of data (Brieman et al. 1984) and such models have shown to capture the spatially varying rules of landscape change (McDonald and Urban 2006). Vegetation type was selected as the first predictor variable, with predicted biomass of 15.5 Mg/500 m² for mixed evergreen and 31.5 Mg/500 m² for redwood-tanoak forests (Fig 3). Such large difference in biomass accumulation by vegetation community type is not surprising considering that redwoods are dominantly large trees in the Big Sur. The models showed positive effect of TMI and PSR on biomass, which reflects the role of energy on plant growth and biomass accumulation (Chen et al. 1999). The ecological effects of distance to non-forest were observed for plots in redwood-tanoak forests, whereby plots that are close to the forest edge (i.e., lower distance to non-forest) had higher biomass compared to plots that were towards inside the forests. The effects of distance to non-forest on biomass accumulation is mediated through their effects on forest microclimate by affecting incoming sunlight and temperature gradient from outer to inner forests (Chen et al. 1999; VanWalleghem and Meentemeyer 2009). Such ecologically mediated effects of forest edge on biomass is of growing interest considering the increasing rates of forest fragmentation associated with increasing deforestation globally (Asbjornsen et al. 2004). Surprisingly, disturbance history did not affect biomass which is likely because many dominant trees in the Big Sur are long lived and persist throughout forest succession. Overall, redwood-tanoak forests that have high soil moisture or are in forest edges (short distance to non forest) are hotspots for biomass accumulation in the Big Sur, and such areas are of interest to enhance terrestrial carbon storage. With smaller sample size, fewer variables were identified as predictors of biomass (Table 3), which is expected in landscape and regional scale studies because smaller sample sizes may not be representative of the wide range in ecological and physiographic variables that influence plant growth and biomass accumulation. However, selection of the most important predictor variables (vegetation type, TMI, PSR) and the model fit were relatively insensitive to the sample size. Thus, biomass prediction models developed with small sample sizes are able to map spatial patterns of biomass distribution resulting from the variation in major predictor variables. The landscape context variation in predictor variables and their predictive relationships with biomass determines biomass spatial distribution across the Big Sur.

We observed that deterministic spatial variation in biomass distribution within spatial scales of ~3.1 km, as observed in this study, could be accounted for by spatial variation in ecological and physiographic variables. Spatial variation in ecological and physiographic variables should therefore be incorporated into biomass mapping models. This is especially relevant for mapping biomass distribution across heterogeneous landscapes where biomass spatial variation occurs within spatial scales that are too fine to be identified by sparsely sampled data. Biomass map derived using RK model (Fig 4) shows spatially heterogeneous pattern of biomass distribution across the Big Sur which would not be captured by univariate kriging model that generates

smooth maps. Our ability to infer and map heterogeneous spatial pattern across the landscape results from RT model which explored the predictive relationship of biomass with ecological and physiographic variables. Based on independent validation, RT model explained 26% of biomass variation with a prediction RMSE of 15.8. When autocorrelated spatial variation was combined with predictions from environmental correlation model, RK model explained larger portion of the variation (41%) and produced more accurate predictions with a RMSE of 13.2. Regional scale analyses have also shown that incorporating covariates like vegetation, elevation difference and soil texture into biomass mapping model leads to improved biomass estimations (Sales et al. 2007). When regression residuals are spatially autocorrelated, regression based predictions combined with spatial prediction of regression residuals have shown to improve predictions (Hengl et al. 2004). Thus, mapping models developed for spatial prediction and mapping of biomass distribution across heterogeneous landscapes, such as the Big Sur, should be based on spatial autocorrelation and predictive association with environmental variables.

Predictive models and their performance can be sensitive to sample size and information contained within the samples; as such, landscape and regional scale predictive models often strike a balance between sample size and prediction performance. We observed that biomass spatial autocorrelation was short ranged with approximately 56% of the spatial variance occurring within 400 m. Thus, mapping biomass distribution based on univariate kriging approach, which relies on sample data and the model of spatial autocorrelation model, would require high density samples to map spatially heterogeneous variation in biomass distribution across the Big Sur. RT models developed with sample size as smaller as 84 identified major predictors of biomass and the model fit remained similar. However, sample sizes smaller than 112 failed to identify spatial autocorrelation structure indicating the limitation of sample size for semivariogram modeling. A minimum of 100 sample size has been recommended to accurately derive a semivariogram (Webster and Oliver 1992). Therefore, samples size can be a limiting factor for spatially-explicit prediction of biomass distribution across heterogeneous landscapes, such as the Big Sur, and we recommend that the sample size be larger than 112.

Assessment of spatial variation in forest biomass distribution in heterogeneous landscapes is challenging, but is needed to improve regional-scale assessments of carbon dynamics. As environmental variables can be autocorrelated at multiple spatial scales, which result largely from environmental forcing or community processes (Legendre 1993; Overmars et al. 2003), our ability to assess their spatial variation is highly influenced by the data density and spatial configuration of field observations. Across the heterogeneous landscapes of Big Sur, our relatively dense network of field observations made it possible to detect localized spatial autocorrelation in biomass distribution. Therefore, landscape- to regional-scale models of forest biomass distribution should incorporate spatial autocorrelation in vegetation and spatial variation in ecological and physiographic factors using data that sufficiently captures the region's landscape het-

erogeneity.

Acknowledgement

The authors gratefully acknowledge financial support from the National Science Foundation (EF-0622770 and EF-0622677), the USDA Forest Service – Pacific Southwest Research Station, and the Gordon & Betty Moore Foundation. The paper was improved through constructive comments from J. Ohlmann, D. Shoemaker, and K. Frangioso. We thank numerous people who provided field support, including: R. Cobb, K. Frangioso, E. Paddock, J. Vieregge, L. Waks, and A. Wickland.

References

- Asbjornsen H, Ashton MS, Vogt DJ, Palacios S. 2004. Effects of habitat fragmentation on the buffering capacity of edge environments in a seasonally dry tropical oak forest ecosystem in Oaxaca, Mexico. *Agricultural Ecosystems and Environment*, **103**: 481–495.
- Anderson LO, Malhi Y, Ladle RJ, Aragão LEOC, Shimabukuro Y, Phillips OL, Baker T, Costa ACL, Espejo JS, Higuchi N, Laurance WF, López-González G, Monteagudo A, Núñez-Vargas P, Peacock J, Quesada CA, Almeida S, Vásquez R. 2009. Influence of landscape heterogeneity on spatial patterns of wood productivity, wood specific density and above ground biomass in Amazonia. *Biogeosciences Discussions*, **6**: 2039–2083.
- Baccini A, Friedl MA, Woodcock CE, Warbington R. 2004. Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters*, **31**: L10501–4.
- Beyer HL. 2004. Hawth's Analysis Tools for ArcGIS. Available at <http://www.spatial ecology.com/htools>
- Blackard JA, Finco MV, Helmer EH, Holden GR, Hoppus ML, Jacobs DM, Lister AJ, Moisen GG, Nelson MD, Riemann R, Reufenacht B, Salajano D, Weyermann DL, Winterberger KC, Brandeis TJ, Czaplewski RL, McRoberts RE, Patterson PL, Tymico RP. 2008. Mapping US forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment*, **112**: 1658–1677.
- Borchert M, Lopez A, Bauer C, Knowl T. 2004. Field guide to coastal sage scrub and chaparral alliances of Los Padres National Forest. USDA Forest Service, Vallejo, CA.
- Breiman L, Friedman J, Olshen R, Stone C. 1984. *Classification and regression trees*. Pacific Grove, CA: Wadsworth..
- Brown S. 2002. Measuring carbon in forests; current status and future challenges. *Environmental pollution*, **116**: 363–372.
- Brown SL, Schroeder P, Kern JS. 1999. Spatial distribution of biomass in forests of the eastern USA. *Forest Ecology and Management*, **123**: 81–90.
- Brown S, Sathaye J, Cannell M, Kauppi P. 1996. Management of forests for mitigation of greenhouse gas emissions. In R.T Watson, M.C. Zinyowera, and R.H. Moss (eds.), *Climate Change 1995: Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses*. Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge and New York: Cambridge University Press, pp. 775–794.
- Callaway RM, Davis FW. 1993. Vegetation dynamics, fire and the physical environment in coastal central California. *Ecology*, **74**(5): 1567–1578.
- Chen J, Saunders SC, Crow RR, Naiman RJ, Broszofsky KD, Mroz GD, Brookshire BL, Franklin JF. 1999. Microclimate in forest ecosystem and

- landscape ecology. *BioScience*, **49**: 558–288–297.
- Clark DB, Clark DA. 2000. Landscape-scale variation in forest structure and biomass in a tropical rain forest. *Forest Ecology and Management*, **137**: 185–198.
- Daly C, Talyon GH, Gibson WP, Parzybok TW, Johnson GL, Pasteris P. 2001. High-quality spatial climate data sets for the United States and beyond. *Trans. ASAE*, **43**: 1957–1962.
- Davis FW, Borchert MI, Flint A, Meentemeyer RK, Rizzo DM. 2010. Pre-impact forest composition and ongoing tree mortality associated with sudden oak death disease in the Big Sur Region, California. *Forest Ecology and Management*, **259**(12): 2342–2354.
- DeAth G, Fabricius KE. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, **81**(11): 3178–3192.
- Dixon RK, Andrasko KJ, Sussman FG, Lavinson MA, Trexler MC, Vinson TS. 1993. Forest sector carbon offset projects: near-term opportunities to mitigate greenhouse gas emissions. *Water, Air, and Soil Pollution*, **70**: 561–577.
- Dubayah RC. 1994. Modeling a solar radiation topoclimatology for the Rio Grande river basin. *Journal of Vegetation Science*, **5**: 627–640.
- FRAP (Fire Resources and Management Program). 1995. Fire management for California ecosystems. California Department of Forest and Fire Protection. Available at <http://frap.cdf.ca.gov/>
- Freeman EA, Moisen GG. 2007. Evaluating kriging as a tool to improve moderate resolution of forest biomass. *Environmental Monitoring and Assessment*, **128**: 395–410.
- Fried JS, Zhou X. 2008. Forest inventory-based estimation of carbon stocks and flux in California forests in 1990. United States Department of Agriculture Forest Service Technical Report, PNW-GTR 750, Washington, DC.
- Gesch, DB. 2007. The National Elevation Dataset. In Maune, D. (ed.), *Digital Elevation Model Technologies and Applications: The DEM Users Manual, 2nd Edition*: Bethesda, Maryland: American Society for Photogrammetry and Remote Sensing, pp. 99–118.
- Gilbert B, Lowell K. 1997. Forest attributes and spatial autocorrelation and interpolation: effects of alternative sampling locations schemata in the boreal forest. *Landscape and Urban Planning*, **37**: 235–244.
- Goovaerts P. 1997. *Geostatistics for natural resources evaluation*. New York: Oxford University Press.
- Goovaerts P. 2001. Geostatistical modeling of uncertainty in soil science. *Geoderma* **103**: 3–26.
- He HS, Shang BZ, Crow TR., Gustafson EJ, Shifley SR. 2004. Simulating forest fuel and fire risk dynamics across landscapes – LANDIS fuel module design. *Ecological Modelling*, **180**: 135–151.
- Hengl T, Heuvelink GBM, Stein A. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, **120**: 75–93.
- Henson P, Usner DJ. 1996. *The natural history of Big Sur*. University of California Press.
- Houghton RA. 1999. The annual net flux of carbon to the atmosphere from changes in land use 1850–1990. *Tellus*, **51B**: 298–313.
- Houghton RA. 2005. Aboveground forest biomass and the global carbon balance. *Global Change Biology*, **11**: 945–958.
- Hu H, Wang GG. 2008. Changes in forest biomass carbon storage in the South Carolina Piedmont between 1936 and 2005. *Forest Ecology and Management*, **255**: 1400–1408.
- Jenkins JC, Chojnacky DC, Heath LS, Birdsey RA. 2003. National-scale biomass estimators for United States tree species. *Forest Science*, **49**(1): 12–35.
- Klaasen W, van Bruegel PB, Moors EJ, Nieveen JP. 2002. Increased heat fluxes near a forest edge. *Theoretical and Applied Climatology*, **72**: 231–243.
- Lamsal S, Grunwald S, Bruland GL, Bliss CM, Comerford NB. 2006. Regional hybrid geospatial modeling of soil nitrate-nitrogen in the Santa Fe River Watershed. *Geoderma*, **135**: 233–247.
- Legendre P. 1993. Spatial autocorrelation: trouble or new paradigm. *Ecology* **74**(6): 1659–1673.
- Lovett GM, Jones CG, Turner MG, Weathers KC. 2005. *Ecosystem functions in heterogeneous landscapes*. New York, NY: Springer.
- Maloney PE, Lynch SC, Kane SF, Jensen CE, Rizzo DM. 2005. Establishment of an emerging generalist pathogen in redwood forest communities. *Journal of Ecology*, **93**: 899–905.
- McDonald RI, Urban DL. 2006. Spatially varying rules of landscape change: lessons from a case study. *Landscape and Urban Planning*, **74**: 7–20.
- Meentemeyer RK, Moody A. 2002. Distribution of plant life history types in California chaparral: the role of topographically-determined drought severity. *Journal of Vegetation Science*, **13**(1): 67–78.
- Meentemeyer RK, Rank NE, Shoemaker DA, Oneal CB, Wickland AC, Frangioso, KM, Rizzo, DM. 2008. Impact of sudden oak death on tree mortality in the Big Sur ecoregion of California. *Biological Invasions*, **10**: 1243–1255.
- Meng Q, Cieszewski C, Madden M. 2007. Large area forest inventory using Landsat ETM+: a geostatistical approach. *Journal of Photogrammetry and Remote Sensing*, **64**: 27–36.
- Mickler RA, Earnhardt TS, Moore JA. 2002. Regional estimation of current and future forest biomass. *Environmental Pollution*, **116**: S7–S16.
- Moore ID, Grayson RB, Ladson, AR. 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, **5**: 3–30.
- Nanos N, González-Martínez SC, Bravo F. 2004. Studying within-stand structure and dynamics with geostatistical and molecular marker tools. *Forest Ecology and Management*, **189**: 223–240.
- Návar J. 2008. Allometric equations for tree species and carbon stocks for forests of northwestern Mexico. *Forest Ecology and Management*, **257**: 427–434.
- Ohmann JL, Crecory MJ, Spies TA. 2007. Influence of environment, disturbance, and ownership on forest vegetation of coastal Oregon. *Ecological Applications*, **17**: 18–33.
- Overmars KP, de Koning GHJ, Veldkamp A. 2003. Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, **164**: 257–270.
- Poulos HM. 2009. Mapping fuels in the Chihuahuan desert borderlands using remote sensing, geographic information systems, and biophysical modeling. *Canadian Journal of Forest Research*, **39**: 1917–1927.
- Poulos HM, Camp AE, Gatewood RG, Loomis L. 2007. A hierarchical approach for scaling forest inventory and fuels data from local to landscape scales in the Davis Mountains, Texas, USA. *Forest Ecology and Management*, **244**: 1–15.
- Saatchi S, Halligan K, Despain DG, Crabtree RL. 2007. Estimation of forest fuel load from radar remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(6): 1726–1740.
- Saatchi S, Malhi Y, Zutta B, Buermann W, Anderson LO, Araujo AM, Phillips OL, Peacock J, ter Steege H, Lopez Gonzalez G, Baker T, Arroyo L, Almeida S, Higuchi N, Killeen T, Monteagudo A, Neill D, Pitman N, Prieto A, Salomão R, Silva N, Vásquez Martínez R, Laurance W, Ramírez HA.

2009. Mapping landscape scale variations of forest structure, biomass, and productivity in Amazonia. *Biogeosciences Discuss*, **6**: 5461–5505.
- Sales MH, Souza CM, Kyriakidis PC, Roberts DA, Vidal E. 2007. Improving spatial distribution estimation of forest biomass with Geostatistics: a case study of Roraima, Brazil. *Ecological Modelling*, **205**: 221–230.
- Schlamadinger B, Marland G. 1996. The role of forest and bioenergy strategies in the global carbon cycle. *Biomass and Energy*, **10**: 275–300.
- Scholz RW, Schnabel U. 2006. Decision making uncertainty in case of soil remediation. *Journal of Environmental Management*, **80**: 132–147.
- Schreuder HT, Gregoire TG, Wood GB. 1993. *Sampling methods for multiresource forest inventory*. New York, NY: John Wiley and Sons, Inc.
- Smith WB. 2002. Forest inventory and analysis: a national inventory and monitoring program. *Environmental Pollution*, **116**: S233–S242.
- Vanwallendael T, Meentemeyer RK. 2009. Predicting forest microclimate in heterogeneous landscape. *Ecosystems*, **12**(7): 1158–1172.
- Webster R, Oliver MA. 2007. *Geostatistics for Environmental Scientists*. Chichester, England: John Wiley and Sons.